Wildfire Impacts on Ozone on June 21, 2015 at the El Paso UTEP Monitoring Site

Dr. Dan Jaffe

May 17, 2017

Executive Summary

Trajectories and CAMX modeling demonstrate that smoke from wildfires in Arizona and New Mexico was transported to El Paso Texas on June 21, 2015. The CAMx modeled arrival time coincides nearly exactly with the time of enhanced PM and ozone (O_3) at the University of Texas at El Paso (UTEP) monitoring site. The PM and O_3 peaks, which appear on June 21, 2015 between 10 am and 2 pm, are present in a ratio that is consistent with published observations of O_3 in aged wildfire plumes. The observed MDA8 on this date was 77 ppb. We used a Generalized Additive Model to estimate the contribution from the fires to the 8-hour O_3 concentration for the UTEP site for June 21, 2015. Our best estimate of the wildfire contribution to the MDA8 on this day is 23 ppb. Applying a 97.5th percentile error bounds based on the EPA guidance method, we calculate a minimum contribution due to the wildfires of 7 ppb to the MDA8.

1. Introduction

This report supplements the information in the TCEQ "EXCEPTIONAL EVENT DEMONSTRATION PACKAGE For the El Paso County Maintenance Area" and the CAMx modeling package, completed by Ramboll-Environ in February 2017. Combined, the TCEQ and Ramboll-Environ documents show clear evidence of wildfire smoke transport to the El Paso UTEP monitoring site on June 21, 2015. In particular, the CAMx modeling shows enhanced smoke from the fires at exactly the time when PM and O_3 are enhanced in El Paso on June 21, 2015. In this report, we demonstrate that the PM and O_3 enhancement ratios are consistent with an aged wildfire source. We use a statistical modeling approach to estimate a wildfire contribution to the MDA8 on June 21, 2015 of 23 ppb. . . We further calculate a minimum contribution of 7 ppb based on the "error bounds" from the EPA guidance document on wildfire exceptional events.

2. Introduction to using statistical models for Exceptional Event (EE) analysis

A number of methods have been used to investigate the impacts of meteorological variables on O_3 concentrations. Camalier (2007) summarizes prior studies on this and developed an approach using Generalized Linear Models. In a project for TCEQ, Alvarado et al (2015) used Generalized Additive Models (GAMs) to investigate the relationship between O_3 and meteorology for 6 six cities in Texas. Jaffe et al (2003) was the first to use the "residual" approach to quantify the amount of O_3 due to wildfires. This approach was further developed and compared against Eulerian models in Jaffe et al (2013). The California Air Resources Board applied this method in a successful exceptional events case demonstration for 2008 California wildfires (CARB 2011), and EPA cited this element in its approval documentation (April 13, 2011). The CARB analysis did not apply the stringent error bound requirements (described below), but was accepted in any case. The approach taken for the present analysis builds on the knowledge from these prior studies.

It is important to note that a statistical model uses observations to identify the usual relationships between O_3 and meteorology. High residuals, the difference between the model prediction and the observation, suggest an unusual or additional source of O_3. However, a statistical model alone cannot identify the cause for a high residual. Possible causes for a significant residual include unusual emissions (e.g. an industrial upset), a stratospheric intrusion or a contribution from a wildfire.

EPA cites the use of statistical regression models in "Guidance on the Preparation of Exceptional Events Demonstration for Wildfire Events that May Influence Ozone Concentrations," dated September 15, 2016, as one of three methods to show that wildfire emissions caused an O_3 exceedance, stating, "the difference between the predictions and observations can provide a reasonable estimate of the air pollution caused by event-related emissions (e.g., emissions from wildfires) provided the analysis accounts for the typical remaining variance of typical days (variability in monitored data not predicted by the model)." (pages 27-28). Our analysis is consistent with the EPA guidance in all respects.

We note that the EPA guidance document also discusses an analysis called "Q/D" (source emissions divided by distance from the fire). We choose not to apply this method as it appears inconsistent with peer reviewed scientific analyses that clearly demonstrate that for most wildfire plumes, O_3 concentrations increase with distance from the fire (Jaffe and Wigder 2012). It is also largely based on Eulerian modeling, which is known to have significant challenges in accurately modeling wildfire plumes (e.g. Baker et al 2016).

3. Observations of fire and smoke on June 20 and 21, 2015

The NOAA Hazard Mapping System (HMS)- Fire and Smoke Product (FSP) for June 20 and 21, 2015 are shown in Figure 1. The HMS product indicates "smoke" in three qualitative ranges of light, medium and heavy. Light smoke is estimated to be in the range of 5-16 μg/m^3 (Brey and Fischer 2016). Note that the HMS products are usually only available once per day, at irregular intervals, so the exact location of the smoke plume at any other point during the day is uncertain. The HMS-FSP can also be obscured by clouds or dust. A narrow or rapidly moving smoke plume is not likely to be depicted accurately by the HMS product. Nonetheless, the HMS-FSP shows substantial transport of smoke to the south and east of these fires on June 20 and 21, consistent with trajectories presented in the original exceptional event documentation. The trajectories indicate transport times of more than one day, depending on which fire is most important.

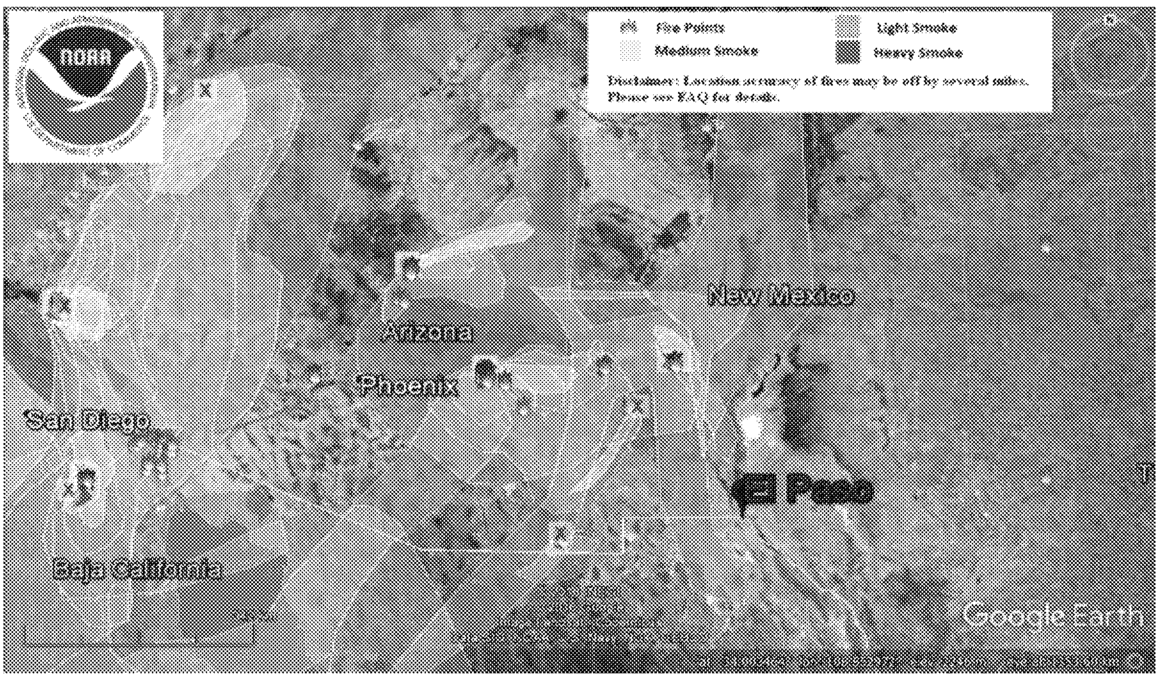


Figure 1. NOAA Hazard Mapping System (HMS)-Fire and Smoke Product for June 20 and 21, 2015

4. Overview of observations on June 21, 2015

Hourly O_3 and PM data for June 21, 2015 are shown in Figure 2. Also shown in Figure 2 is the average diurnal cycle for June and July 2015, combined. The typical pattern for O_3 is a slow daytime buildup, peaking around 2 pm (1400), whereas PM usually peaks at night. During the daytime, PM usually reaches its minimum values due to a rising boundary layer. However, on June 21, 2015, the pattern is rather different for both PM and O_3 with a rapid rise starting at 10 am. On this day, both PM and O_3 peak at 12 noon. The O_3 rise between 10 am and 12 noon is 46 ppb, which is too fast to be explained by a typical urban photochemical buildup. For example median O_3 production rates for Houston were only 10 ppb per hour for September 2013 (Mazucca et al 2016). The rapid increase of both PM and O_3 are suggestive of a narrow plume that swept thru the region. By 2 pm, near the time when O_3 usually peaks, it has rapidly declined, as did PM. Between the hours of 11 and 6pm, there is an excellent correlation between O_3 and $PM_{2.5}$, with a slope of 6.1 and an $R^2 = 0.96$. We also note that TCEQ has reported no evidence for unusual emissions in the El Paso region for this date (personal communication from Erik Gribbin, Technical Specialist, TCEQ, May 15, 2017).

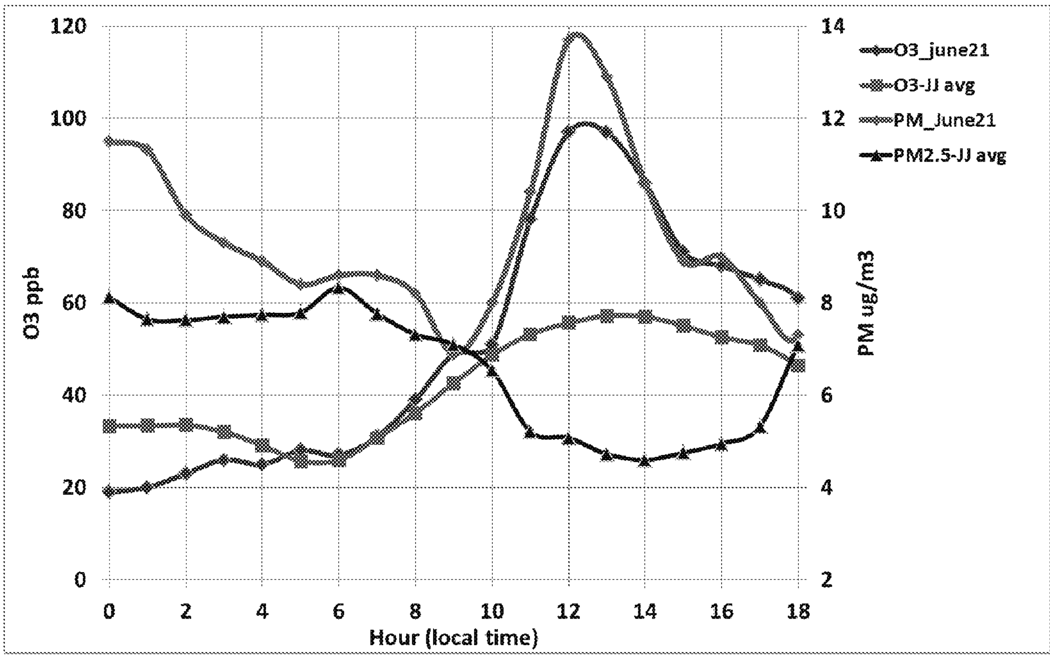


Figure 2. Hourly O_3 and $PM_{2.5}$ data for June 21, 2015 for the UTEP monitor along with averaged hourly data for all days in June and July, 2015 (O3-JJ avg and $PM_{2.5}$-JJ avg).

Using the hourly data, we can estimate the consistency between the PM and O_3 levels. To do this, we calculate ΔPM and ΔO_3 value as the difference between the June-July, 2015 monthly average for each hour and the observations on June 21, 2015. Figure 3 shows a graphical representation of this method. From this we get ΔPM and ΔO_3 values of 8.6 $\mu g/m^3$ and 41.4 ppb, respectively, at noon when the plume passed over. This gives an enhancement ratio of 4.8 ppb of O_3 per $\mu g/m^3$ of $PM_{2.5}$.

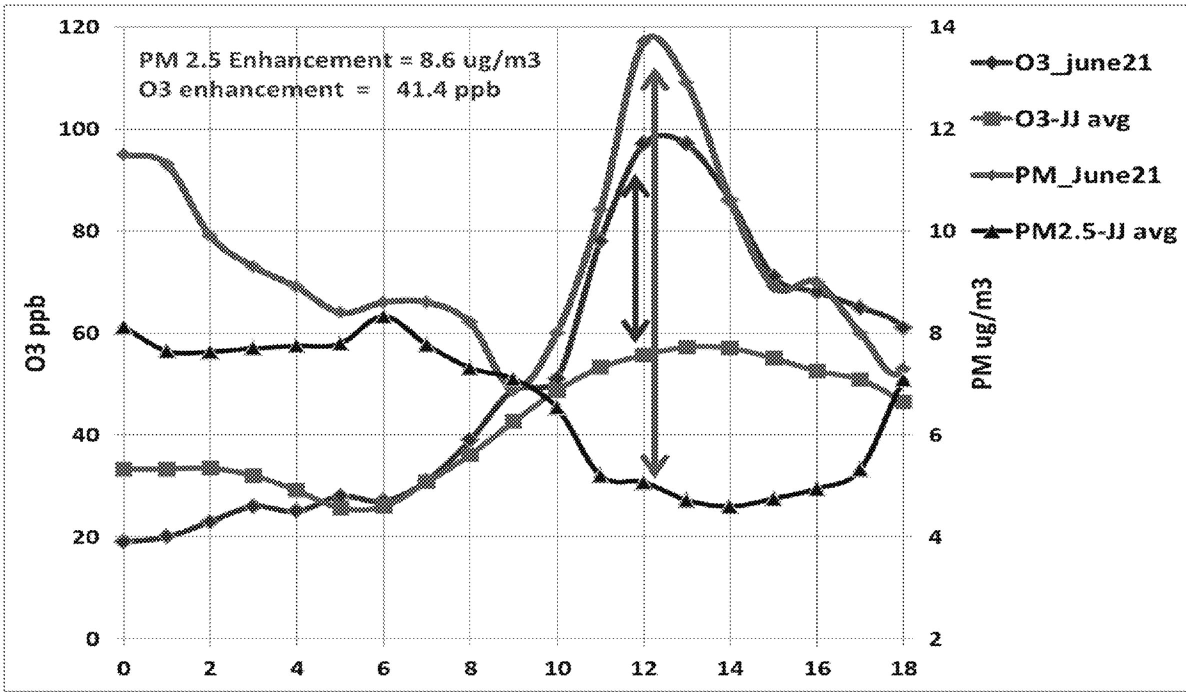


Figure 3. Hourly O_3 and $PM_{2.5}$ data for June 21, 2015 for the UTEP monitor along with averaged hourly data for all days in June and July, 2015 (O3-JJ avg and $PM_{2.5}$-JJ avg). This is the same as Figure 2, with the addition of arrows representing the calculated $\Delta PM_{2.5}$ (brown arrow) and ΔO_3 (red arrow) values. The enhanced $PM_{2.5}$ and O_3 values (ΔPM and ΔO_3) for June 21 at noon are 8.6 $\mu g/m^3$ and 41.4 ppb, respectively, leading to an $\Delta O_3/\Delta PM_{2.5}$ enhancement ratio of 4.8 ppb of O_3 per $\mu g/m^3$ of $PM_{2.5}$.

We can compare this value to literature values for aged wildfire plumes and under the assumption that PM has not been lost during transit (e.g. clouds or precipitation). Note that the chemistry of O_3 production in a wildfire is highly complex and depends on many factors including emissions, processing and photochemistry. It is important to recognize that generally PM will **decrease** with distance from a fire, while O_3 will **increase** (Jaffe and Wigder 2012). So we do not expect a simple linear relationship between O_3 and PM. Nonetheless we can compare our PM-O_3 relationship to what is seen in the literature for wildfire plumes transported more than one day. In many cases, previous studies report $\Delta PM_{2.5}/\Delta CO$ and/or $\Delta O_3/\Delta CO$ from wildfire smoke plumes. Therefore we can estimate the $\Delta O_3/\Delta PM_{2.5}$ ratio from the following relationship:

$$\frac{\Delta O3}{\Delta PM} = * \frac{\Delta O3}{\Delta CO} * \frac{\Delta CO}{\Delta PM}$$

Laing et al (2017) reports on the PM to CO relationship for 25 wildfire events as seen in 8 urban locations in the Western U.S. They report an average **$\Delta PM_{2.5}/\Delta CO$** ratio in wildfire smoke of 0.13 µg/m^3 per ppb of CO with a range of 0.06-0.23, relatively consistent with the emission factors reported by Akagi et al (2011), which are in the range of 0.10 to 0.20 µg/m^3 per ppb. These correspond to **$\Delta CO/ \Delta PM_{2.5}$** ratios of 7.7, 16.7 and 4.3 ppb/µg/m^3, respectively. This paper, although still in review for publication in a peer-reviewed journal, has been requested by EPA for inclusion in a technical support document (A.Mebust, U.S. EPA Region 9, Air Quality Analysis Office, personal communication, March 30, 2017).

For $\Delta O_3/\Delta CO$, we use a published review of more than 100 peer reviewed studies on wildfire smoke. Jaffe and Wigder (2012) report mean $\Delta O_3/\Delta CO$ ratios for sub-tropical wildfire plumes aged 2-5 days of 0.35 ppb/ppb, with a range of 0.26-0.42 ppb/ppb. Thus we can calculate a mean, maximum and minimum value for the $\Delta PM/\Delta O_3$ enhancement ratio for wildfire plumes aged 2-5 days:

Mean value:

$$\frac{\Delta O3}{\Delta PM} = 0.35 * 7.7 = 2.7 \frac{ppb}{ug\ per\ m3}$$

Maximum value:

$$\frac{\Delta O3}{\Delta PM} = 0.42 * 16.7 = 7.0 \frac{ppb}{ug\ per\ m3}$$

Minimum value:

$$\frac{\Delta O3}{\Delta PM} = 0.26 * 4.3 = 1.1 \frac{ppb}{ug\ per\ m3}$$

The June 21, 2015 enhancement ratio $\Delta O_3/\Delta PM_{2.5}$ enhancement ratio, 4.8 ppb of O_3 per µg/m^3 of $PM_{2.5}$, fits into the middle the range based on previously published data.

So, in summary, the observations indicate a rapid increase in both $PM_{2.5}$ and O_3 at 10 am on June 21, 2015 at the same time as the CAMx model shows transport of emissions from the wildfires. The timing of the CAMx modeling coincides nearly exactly with the time of enhanced $PM_{2.5}$ and O_3 in El Paso. The rate of increase in O_3 and $PM_{2.5}$ is inconsistent with typical urban photochemistry, but clearly indicates the presence of a narrow plume of PM and O_3 that came thru El Paso. The $\Delta O_3/\Delta PM_{2.5}$ enhancement ratio of 4.8 ppb of O_3 per $\mu g/m^3$ of $PM_{2.5}$ is consistent with published studies indicating a range of 1.1-7.0 ppb O_3 per $\mu g/m^3$ of $PM_{2.5}$ for wildfire smoke plumes that have been transported 2-5 days. We next turn to a quantification of the wildfire effects on O_3.

As a further examination of the PM-O_3 relationship on June 21, 2015, we evaluated this relationship for every day in 2010-2015 with an MDA8 greater than 70 ppb. Out of these 30 days, June 21, 2015 is the only day with a statistically significant positive correlation between PM and O_3 (shown in Table A1). This further confirms the importance of smoke on June 21, 2015.

5. Contributions from fires to the MDA8 on June 21, 2015 using a Generalized Additive Model

To calculate the contribution from the fires to the MDA8, we developed a Generalized Additive Model for hourly O_3 concentrations for the El Paso UTEP site. The model used hourly O_3 data from June-July 2010-2015. For the final model, we used meteorological data from the National Climate Data Center (NCDC) and back-trajectories calculated using the NOAA Hysplit model at an arrival height of 500 meter above ground level. We note that using NCDC data is consistent with EPA practices. We also examined meteorological data from the EPA AirnowTech site, but found that it gave no improvement in the results, compared to the NCDC met data. we modeled the hourly O_3 concentration using the GAM "mgcv" package in R software. We initially examined 81 daily variables including day of year, weekday/weekend, Hysplit back trajectory variables and meteorological variables. Because many of the variables overlap and some have large amounts of missing data, we reduced this list to a much smaller number of variables to build the El Paso GAM model. Also, modeling of the hourly data limited the numbers of variables that could be incorporated in a reasonable amount of computational time. A few additional variables were excluded as they did not have data for June 21, 2015. The final/best model was developed from 15 numerical variables and 3 categorical variables. Table 1 shows the list of final variables used in the model.

Each parameter was fit with a spline function to incorporate its influence on O_3 into the model. The final model results were evaluated based on the Akaike Information Criterion

(AIC), the overall model fit (R^2), by evaluating the relationship of the residuals to individual variables and the statistical significance of each individual parameter. The AIC is a widely used evaluation tool for GAM (Wood 2006) that considers not only the model fit, but also the possibility of over-fitting by inclusion of additional predictors that have minimal effect on the model. In general, the model with the lowest AIC is preferred (Wood 2006; Alvarado 2015). We found that a log-link with a Gaussian error functions gave the best results (best R^2 and lowest residuals), similar to the results of Camalier et al (2007).

Table 1. Parameters used in the General Additive Model for El Paso O_3 data.

Variable	Description	Source	Type C or N[1]	Variable name
TR16Q	See below[2]	NOAA Hysplit	C	TR16Q
Month	Month	Month	C	Month
WD1017	Vector averaged wind direction for Hrs 0-17	NCDC	C	WD1017
Year	Year	Year	N	V1
DOY	Day of Year	Day of Year	N	V2
TR16D	See below[3]	NOAA Hysplit	N	V3
Hr	Hour of day	Hour of day	N	V4
TMAX	Daily max temperature	NCDC	N	V5
TAVG	Daily average temperature	NCDC	N	V6
TMIN_PREV_Night	Min temperature previous night	NCDC	N	V7
DPAVG	Daily average dew point	NCDC	N	V8
DPMAX017	Daily maximum dew point for hours 0-17	NCDC	N	V9
DPMIN017	Daily minimum dew point for hours 0-17	NCDC	N	V10
SLPAVG	Daily average sea level pressure	NCDC	N	V11
SLP017	Daily average sea level pressure for hours 0-17	NCDC	N	V12
WS017	Vector averaged wind speed for Hrs 0-17	NCDC	N	V13
WS617	Vector averaged wind speed for Hrs 6-17	NCDC	N	V14
WS1017	Vector averaged wind speed for Hrs 10-17	NCDC	N	V15

[1]C denotes categorical variables, N denotes numerical variables.

[2] TR16Q represents the Hysplit backward trajectory quadrant after 24 hour travel for trajectory initialized at 16 GMT (1000 LT). Trajectories arriving at 20 GMT were also evaluated, but gave weaker results.

[3] TR16D represents the Hysplit backward trajectory 24 hour distance from starting point for trajectory initialized at 16 GMT (1000 LT). Trajectories arriving at 20 GMT were also evaluated, but found to give weaker results.

Figure 4 shows a comparison of the observed hourly and modeled O_3 data. The R^2 is 0.64, with a slope of 1.004 and an intercept of -0.12. At low values, there is a slight over-prediction, but this generally goes away at higher values. This is further demonstrated in Figures 5 and 6. Figure 4 also shows the data point for noon on June 21st, 2015. This point is more than 40 ppb from the fit line. Figure 5 shows a histogram of the model residuals. The mean is near zero (-0.02 ppb) and the standard deviation is 9.2 ppb. Figure 6 shows the model residual as a function of the model prediction. The model predictions are shown in 10 ppb bins. Each bin has a mean near zero, with the exception of the 10 ppb bin, which has only 6 data points.

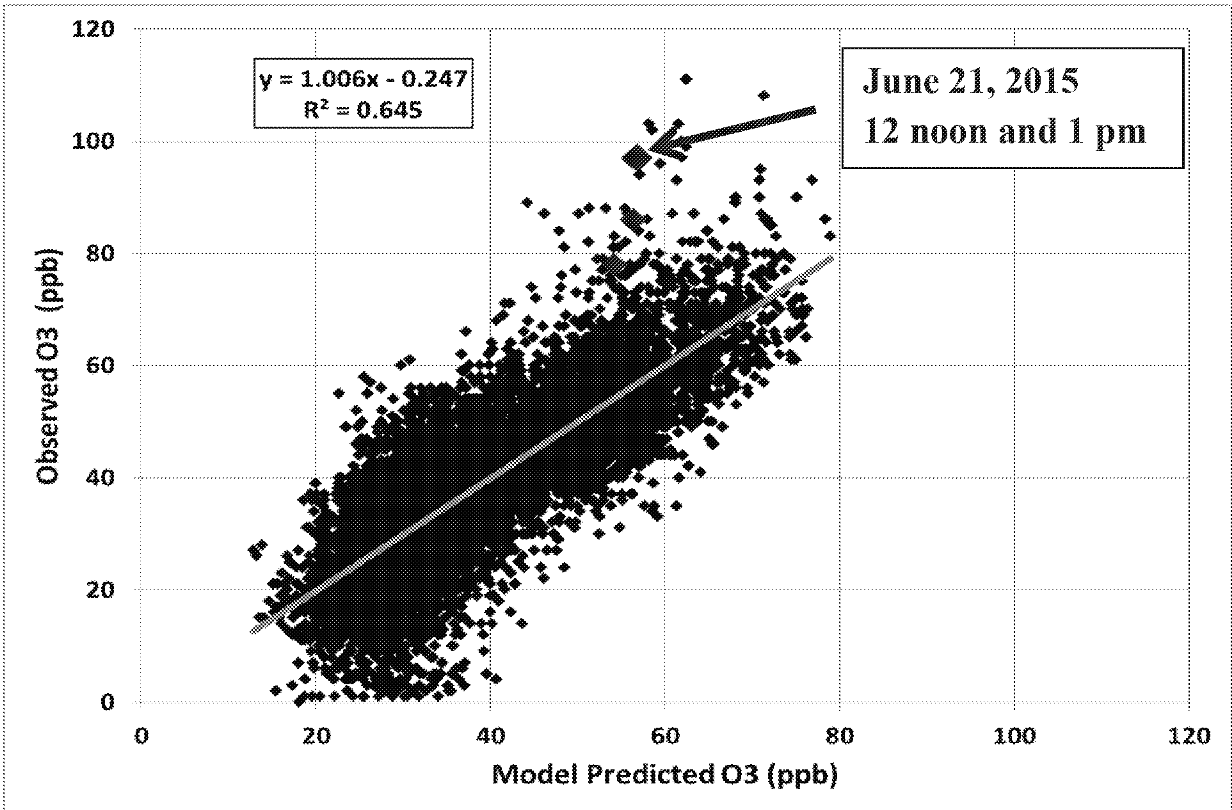


Figure 4. GAM modeled vs observed hourly O_3. A linear regression fit (orange line) between the modeled and observed values yields an $R^2 = 0.64$, a slope of 1.004 and an intercept of -0.12. The data points for June 21, 2015 at 11am – 2pm are highlighted in red. Note that two points overlap (12 noon and 1 pm) as indicated by the larger marker.

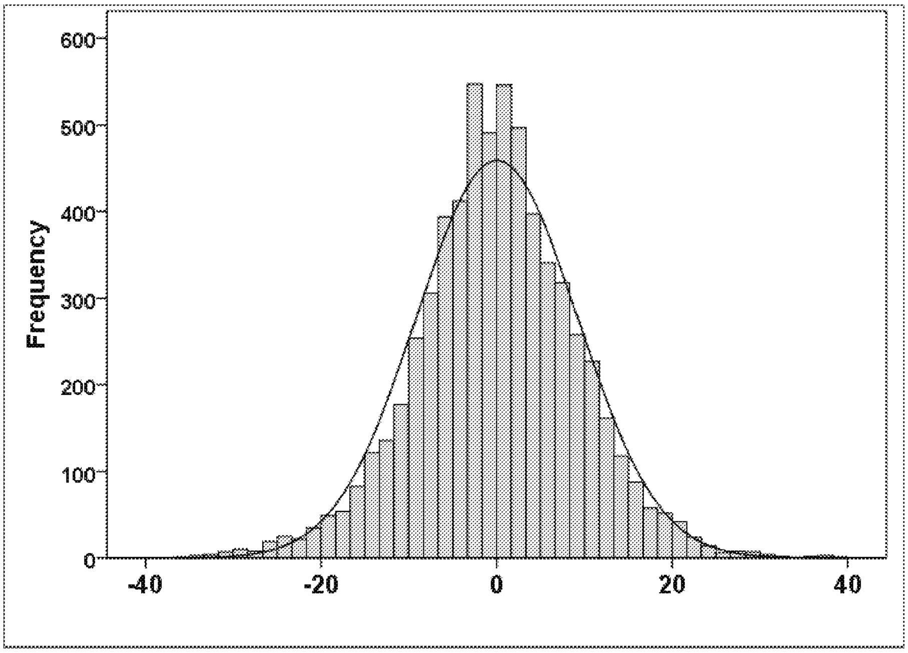


Figure 5. Histogram of the model residuals. The mean residual is -0.02 ppb and the standard deviation is 9.2 ppb. The figure includes 8047 data points.

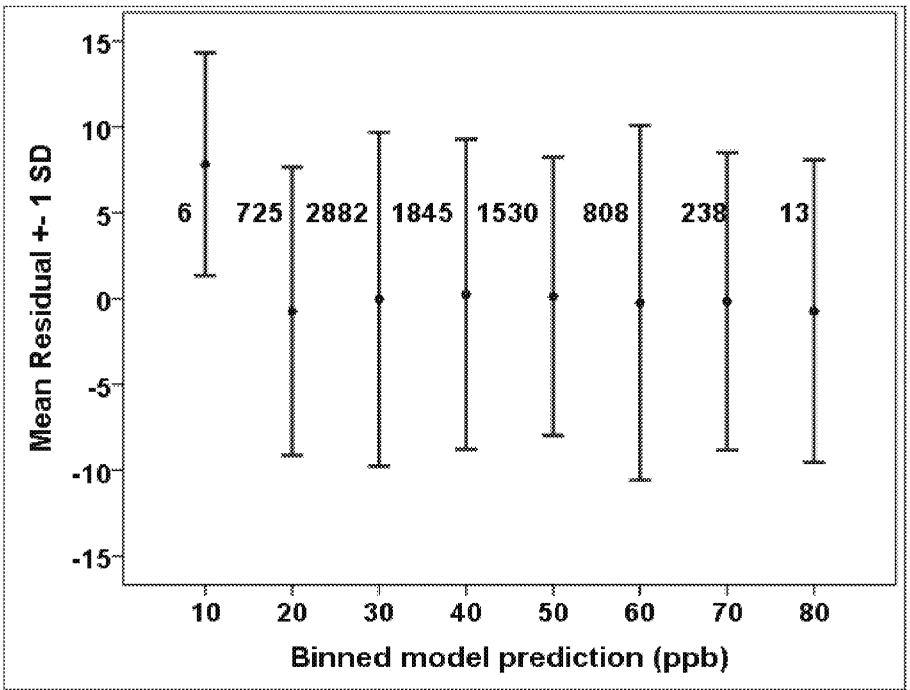


Figure 6. Mean model residual (±1 sd) binned by the model prediction in 10 ppb increments. The bins are centered on 10, 20, 30, 40, 50, 60, 70 and 80 ppb. So, for example, the 60 ppb bin, includes all model predictions between 55-65 ppb. The numbers to the left of each bar indicate the number of data points in that bin. For June 21, 2015 at noon, the model predicts a value of 56 ppb, whereas the observed value was 97 ppb. This yields a residual O_3 amount of 41 ppb in the hourly average. This residual is more than 4 standard deviations away from the mean.

The model residual can be further characterized by its percentile distributions. Table 2 shows the percentile cut-points for all residuals (left side of table) and only the positive residuals right side.

Table 2. Percentile distributions points for residuals of GAM fits. The left 3 columns give the percentiles for all residuals and the right 2 columns give the percentiles for only the positive residuals. Note that the 97.5th percentile of all residuals is equivalent to the 95th percentile of the positive residuals.

Cut point for all residuals					Cut point for positive residuals only	
Percentile	**Value**	**Percentile**	**Value**	***	**Percentile**	**Value**
2.5	-19.2	52.5	0.5	***	5	0.5
5	-15.1	55	1.0	***	10	1.0
7.5	-12.9	57.5	1.5	***	15	1.5
10	-11.0	60	2.0	***	20	2.0
12.5	-9.7	62.5	2.5	***	25	2.5
15	-8.7	65	3.1	***	30	3.0
17.5	-7.8	67.5	3.6	***	35	3.6
20	-6.9	70	4.2	***	40	4.1
22.5	-6.2	72.5	4.9	***	45	4.9
25	-5.6	75	5.7	***	50	5.6
27.5	-4.8	77.5	6.5	***	55	6.5
30	-4.2	80	7.3	***	60	7.3
32.5	-3.6	82.5	8.1	***	65	8.1
35	-3.0	85	9.1	***	70	9.1
37.5	-2.6	87.5	10.2	***	75	10.2
40	-2.1	90	11.4	***	80	11.4
42.5	-1.5	92.5	12.9	***	85	12.9
45	-1.0	95	14.9	***	90	14.9
47.5	-0.5	**97.5**	**18.6**	***	**95**	**18.6**
50	0.0			***		

Additional information on the model and quality control measures are included as an Appendix to this report. This includes probabilities for individual predictors, model cross validation with time periods not included in the model development and an evaluation of the model residuals for only the mid-day time periods time periods. The fact that the predictions are unbiased (Figure 5), that the predictions are unbiased at all predictions levels (Figure 6) and that the predictions are unbiased in the critical mid-day period (see appendix) all indicate that the model is unbiased. We now turn our attention to using the model to quantify the impacts from the wildfires.

Given that the model is unbiased, the best estimate for the MDA8 in the absence of fires for June 21, 2015 is obtained from the modeled hourly values. The modeled hourly values yield an estimated MDA8 of 54 ppb, and therefore a contribution from the fires of 23 ppb (77-54 ppb).

6. Applying the EPA guidance to estimate the minimum contribution to the MDA8

The most likely contribution to the MDA8 from the fires on June 21, 2015 was 23 ppb. To estimate the minimum contribution, we have applied the EPA guidance on use of statistical models. This guidance states *"The difference between the predicted values and the measured values are analyzed, and the 95th percentile of those positive differences (observed O_3 is greater than predicted) is recorded. This 95 percent error bound is added to the O_3 value predicted by the regression equation for the flagged days, <u>and any difference between this sum and the observed O_3 for the flagged day may be considered an estimate of the O_3 contribution from the fire..."</u>* Since the 95th percentile of positive values is equivalent to the 97.5th percentile of all values (see Table 2) we refer to this error limit as the "97.5 percentile error limit". From the GAM fits to the hourly data, we find that the mean, standard deviation, 95th and 97.5th percentiles are: -0.02, 9.2, 14.9 and 18.6 ppb respectively. In this next section, we follow the EPA guidance procedure precisely and show that the wildfire contribution on June 21, 2015, exceeds this 97.5 percentile error value.

Figure 7 shows the observed hourly O_3 and the fit O_3, along with the 97.5th percentile error bounds. Also shown on the figure are the observed hourly $PM_{2.5}$ values. For most of the 24-hour period, the observed O_3 is very well modeled by the GAM fits. Only during the midday period (11 am- 2 pm) do the model values fall outside this 97.5th percentile. <u>This is the exactly the same time when PM is enhanced by smoke.</u> We followed the EPA guidance to compute the impact from the fires by excluding only the contribution to the hourly O_3 values that is above the

97.5th percentile error boundary. Figure 8 shows a representation of the procedure. This results in a reduction of the hourly values (11 am-2 pm) by 5, 21, 21 and 10 ppb, respectively. The MDA8 is recalculated using the observed O_3 for the remainder of the time period. The calculated MDA8 is 70 ppb. Thus we conclude that the wildfires contributed 7 ppb to the MDA8. The 7 ppb is a minimum estimate and only considers the contribution above the 97.5th percentile error bound.

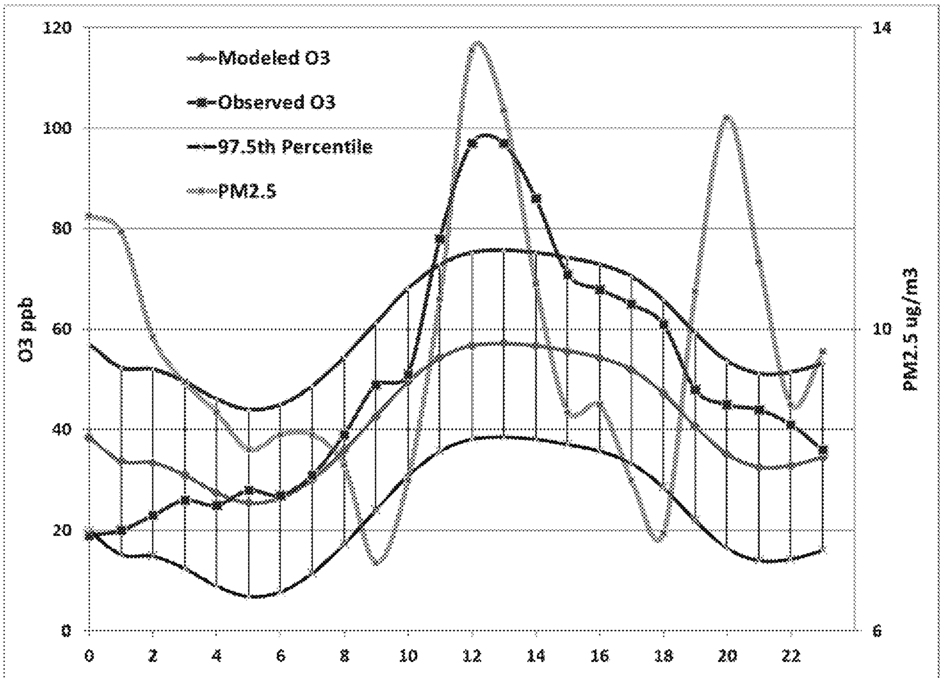


Figure 7. Observed hourly O_3 and the GAM fit O_3, along with the 97.5th percentile error bounds on the GAM fit for June 21, 2015. Also shown is the hourly $PM_{2.5}$ observations.

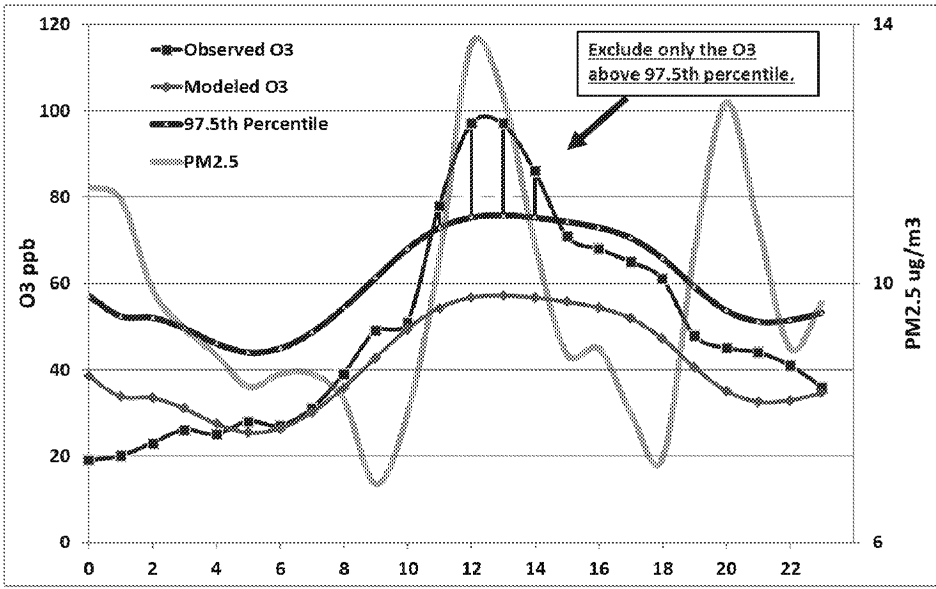


Figure 8. Same as Figure 7, but with the excluded O_3 shown by the red vertical lines. Only the upper 97.5th percentile bounds are shown. Following this procedure yields a contribution of 7 ppb to the MDA8 from the wildfire influenced period.

7. Matching day analysis

A good contrast to this smoke influenced day (June 21, 2015) are the days before and after (June 20 and 22nd, 2015), which had very similar meteorology. All three had relatively high daily max temperatures, similar trajectory distances, sea level pressures and previous night temperatures. Figure 9 shows the observed hourly O_3 and $PM_{2.5}$ data, along with the GAM fit for O_3. Table 3 compares observed and modeled MDA8s and key meteorological parameters. As noted above $PM_{2.5}$ and O_3 are highly correlated on June 21, peaking at noon, whereas on all other days, PM is low during the day and high at night. For June 20 and 22nd, this is understood as being caused by nighttime accumulation of $PM_{2.5}$ in a shallow mixed layer and daytime photochemical production of O_3, at a time when the boundary layer height is increasing. For June 21st, the enhanced peak in O_3 and simultaneous peak in $PM_{2.5}$ are due to the wildfire influence. Figure 9 further shows that the GAM fits to the observations for June 20 and 21 are very good, whereas the observations are much higher than the fit during the daytime on June 21st. So, the matching day analysis leads us to conclude that June 21, 2015 was different from the day before and the day afterwards in multiple respects. This includes the timing of the $PM_{2.5}$ peak and the ability of the GAM fits to accurately reproduce the data, despite having similar meteorological conditions. All of these factors are consistent with O_3 production associated with the transport of smoke.

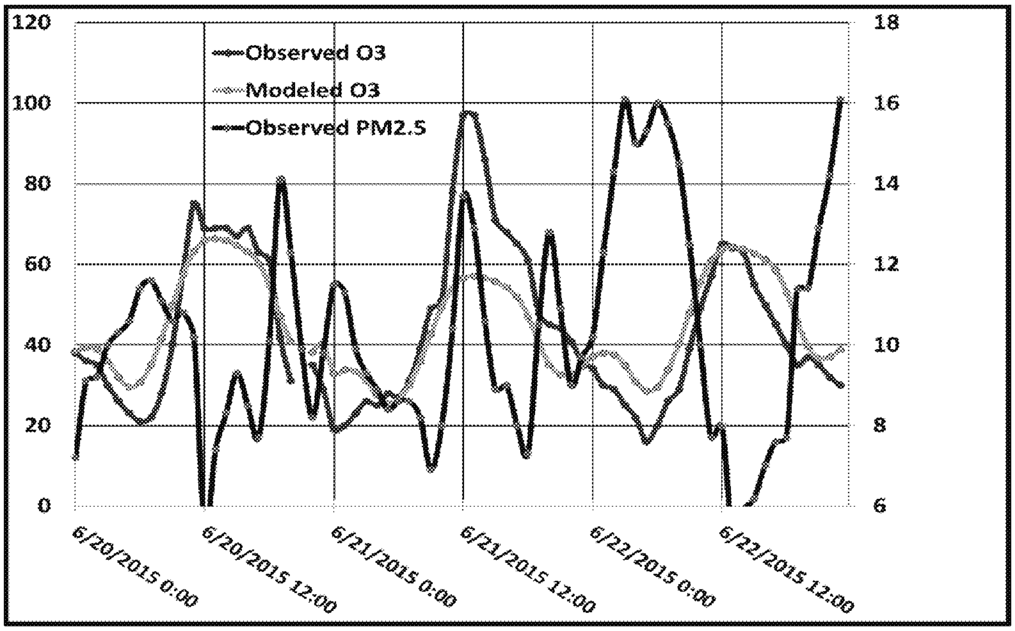


Figure 9. Observed O_3 and $PM_{2.5}$ and modeled O_3 for June 20-22, 2015. The modeled MDA8 values for June 20 and 22 were within 5 ppb, whereas for June 21, the observed MDA8 was 23 ppb higher than the modeled.

Table 3. Comparison of observed O_3 and PM, modeled fits and meteorological parameters for June 20, 21 and 22, 2015.

	6/20/2015	6/21/2015	6/22/2015
Observed MDA8	67	77	56
Modeled MDA8	63	54	61
Obs - Modeled MDA8	+4	23	-5
Trajectory quadrant (after 24 hrs)	SE	SW	SE
Vector averaged wind direction for hours 10-17 (deg)	268	200	129
Trajectory distance (after 24 hours transport, km)	388	372	325
TMAX (F)	103.0	102.0	103.0
TAVG (F)	88.5	90.2	89.2
TMIN previous night (F)	77.0	77.0	75.0
DPAVG (F)	49.7	44.3	51.8
SLP AVG (mbar)	1006.4	1006.4	1007.5
Wind speed between hours 0-17 (kts)	4.0	3.6	3.2
Time of max PM	7 pm	12 pm (noon)	3 am and 11 pm

8. Conclusion

We have evaluated the El Paso UTEP O_3 and $PM_{2.5}$ for June 21, 2015. Both O_3 and $PM_{2.5}$ are elevated in the mid-day (10am-2pm) and in a ratio which is consistent with published data on aged wildfire smoke plumes. The pattern in $PM_{2.5}$ on June 21, 2015 is significantly different from the usual pattern, where PM peaks at night. The CAMx modeling demonstrates transport of smoke from the multiple fires burning in Arizona and New Mexico. Thus, the pattern of a daytime peak in $PM_{2.5}$, the enhancement ratio of $\Delta PM_{2.5}/\Delta O_3$ and the CAMx modeling all support the conclusion that smoke was transported to the UTEP site at mid-day on June 21, 2015. A quantitative assessment of the fire contributions to O_3 was made using a Generalized Additive Model. The model was shown to be unbiased at all prediction levels and for the daytime maximum. Using this model and following the EPA guidance, we calculate that the fires contributed to be 23 ppb to the MDA8 at the UTEP site for June 21, 2015. Separately, we calculated the minimum contribution of 7 ppb to the MDA8 based on the 97.5th percentile using the EPA wildfire guidance method. Finally a matching day analysis was completed using the day before (June 20th) and the day after the smoke event (June 22nd). These three days had very similar meteorology. The model fits for June 20 and 22 were within 5 ppb, whereas for June 21 the model was 23 ppb low. In addition, June 21 was the only day with a midday peak in

$PM_{2.5}$. Thus we conclude that the matching day analysis demonstrates an unusual impact on O_3 for June 21, 2015. Therefore, the preponderance of evidence indicates that an unusual event occurred on June 21. Specifically, the available evidence suggests significant wildfire influence on the El Paso MDA8, contributing at least 7 ppb and most likely 23 ppb to the observed exceedance.

References

Akagi, S.K. et al. (2011) Emission factors for open and domestic biomass burning for use in atmospheric models. Atmos. Chem. Phys. 11, 4039-4072, 10.5194/acp-11-4039-2011,.

Alvardo M., Lonsdale C., Mountain M., and Hegarty J. (2015) Investigating the Impact of Meteorology on O3 and PM2.5 Trends, Background Levels,and NAAQS Exceedances, TCEQ Contract No. 582-15-50415, Final report from Atmospheric and Environmental Research, Inc. Available at: https://www.tceq.texas.gov/assets/public/implementation/air/am/contracts/reports/da/5821554118FY1501-20150831-aer-MeteorologyAndO3PMTrends.pdf

Baker, K.R., Woody, M.C., Tonnesen G.S., Hutzell W., Pye H.O.T., Beaver M.R., G. Pouliot G., and Pierce T. (2016) Contribution of regional-scale fire events to ozone and PM2.5 air quality estimated by photochemical modeling approaches. Atmos. Envir. 140, 539-554.

California Air Resources Board (CARB 2011). Exceptional Events Demonstration for 1-h Ozone Exceedances in the SacramentoRegional Nonattainment Area Due to 2008 Wildfires, Updated Documentation March 30, 2011. Available at: https://www.epa.gov/air-quality-analysis/exceptional-events-submissions-table

Camalier L., Cox W., and Dolwick P. (2007) The effects of meteorology on ozone in urban areas and their use in assessing ozone trends. Atmos. Envir. 41, 7127–7137.

Jaffe, D. et al. (2004) Long-range transport of Siberian biomass burning emissions and impact on surface ozone in western North America. Geophys. Res. Lett. 31, 10.1029/2004gl020093.

Jaffe, D.A., Wigder, N.L. (2012). Ozone production from wildfires: A critical review. Atmos. Environ. 51, 1-10, 10.1016/j.atmosenv.2011.11.063.

Jaffe, D.A., Wigder, N., Downey, N., Pfister G., Boynard A., and Reid S.B. (2013) Impact of wildfires on ozone exceptional events in the western US. Environ. Sci. Technol. 47, 11065-11072, 10.1021/es402164f.

Laing J.R., Jaffe D.A., Slavens A.P., Li W., Wang, W. (2017) Using ΔPM2.5/ΔCO and ΔNOy/ΔCO Enhancement Ratios to Identify Wildfire Smoke in Urban Areas. Submitted to Aerosol and Air Quality Research, Feb. 2017.

Mazucca G.M. et al. (2016) Ozone production and its sensitivity to NOx and VOCs: results from the DISCOVER-AQ field experiment, Houston 2013. Atmos. Chem. Phys., 16, 14463–14474.

U.S. Environmental Protection Agency (2015) Office of Air Quality Planning and Standards Draft Guidance on the Preparation of Exceptional Events Demonstrations for Wildfire Events that May Influence Ozone Concentrations. Available at: https://www.epa.gov/sites/production/files/2015-11/documents/o3_draft_wildfire_guidance.pdf

Wood, S. N. (2006), Generalized Additive Models: An Introduction with R, part of the "Texts in Statistical Science" series, Chapman & Hall/CRC, New York.

Appendix:

Table A1. PM-O_3 Correlation between 10am-5pm (inclusive) on all days in 2010-2015 with MDA8>70. For statistical significance, need R^2 = 0.66 or higher. Only two dates: 6/22/11 and 6/21/15 have significant correlations and only 6/21/15 has a significant <u>positive</u> correlation. These dates are shown in bold below.

Date	MDA8 (ppb)	O_3-$PM_{2.5}$ SLOPE (ppb per ug/m3)	R^2
7/13/10	87	-3.8	0.26
7/3/13	82	-4.2	0.12
7/19/10	81	2.7	0.17
6/17/15	81	3.4	0.20
6/4/11	78	3.4	0.49
6/22/11	**78**	**-1.7**	**0.70**
7/13/12	77	-2.8	0.61
6/21/15	**77**	**7.1**	**0.90**
7/12/12	75	2.1	0.57
5/24/13	75	2.8	0.47
6/11/13	75	0.0	0.00
7/15/14	75	-0.1	0.00
8/20/10	74	-8.0	0.26
8/4/12	74	5.3	0.30
8/31/12	74	1.1	0.00
9/2/12	74	0.8	0.06
8/10/15	74	2.3	0.11
8/10/10	73	-3.2	0.25
4/28/13	73	-0.7	0.06
8/17/13	73	-2.0	0.29
8/19/13	73	1.5	0.02
6/10/14	73	6.5	0.40
7/15/10	72	-2.0	0.06
6/28/12	72	-1.8	0.60
7/14/12	72	4.7	0.52
6/21/14	72	6.0	0.60
6/29/15	72	-1.6	0.01
7/20/11	71	0.7	0.00
6/29/12	71	-1.3	0.13
8/12/12	71	0.6	0.05
8/21/12	71	4.7	0.44

Further GAM details and quality control

The main report describes the basic model development and quality control including:

1. Model fit and R^2;
2. Overall distribution of model residuals;
3. Distribution of residuals vs binned model prediction;
4. Residuals percentile cut-points, including 95^{th} and 97.5^{th} percentiles.

Here we describe some additional details and quality control tests.

1. Model validation when individual years were excluded;
2. Examination of model performance for only the daytime/high O_3 period (10am-2pm).
3. Individual probabilities for each predictor variable.
4. Examination of probabilities associated with each predictor

One step in model validation is comparison of predictions using data that were not part of the original model development. To do this, we recalculated the GAM predictions while excluding one year of data. The model fits were then used to estimate hourly O_3 for the year that was excluded. Shown below are the results for predicted 2015 data, when 2015 data were excluded from the original model computation. The correlation coefficient remains acceptable (R^2=0.56), but with some deterioration as would be expected.

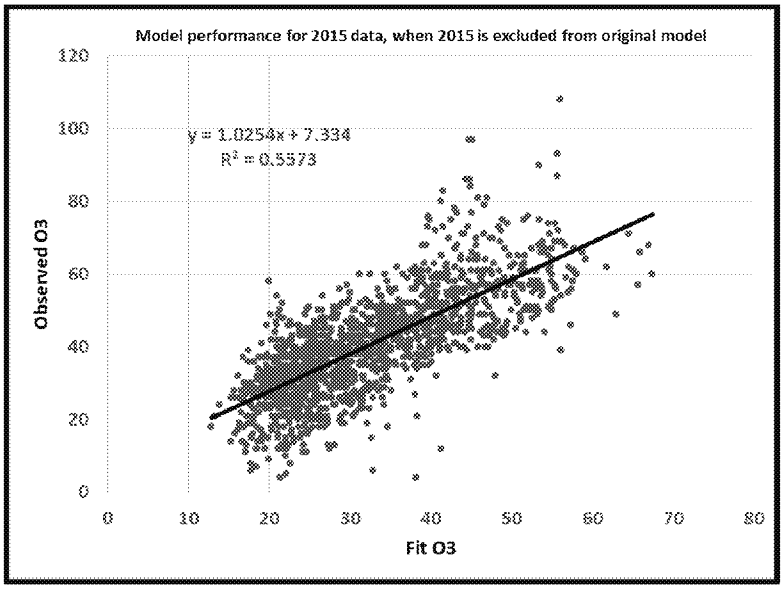


Figure A1: This shows observed vs model calculated O_3 (fit O_3) for 2015, when the 2015 data are excluded from the original model development. So this figure includes 1/5th as much data as shown in Figure

The next check we conducted was to evaluate any possible model bias for the highest O_3 values during the daytime. This is shown in Figure A2. Here we have plotted the observed vs model predicted hourly observations, but only for the hours of 10am-2pm. As we saw earlier (Figure 4), the model is biased high at very low concentrations, but is unbiased at high concentrations. There are a few points that are significantly off the line, including the mid-day points for June 21, 2015. The residuals for this comparison are shown in Figure A3. The standard deviation of these residuals is nearly the same (8.8) as the full dataset (9.2).

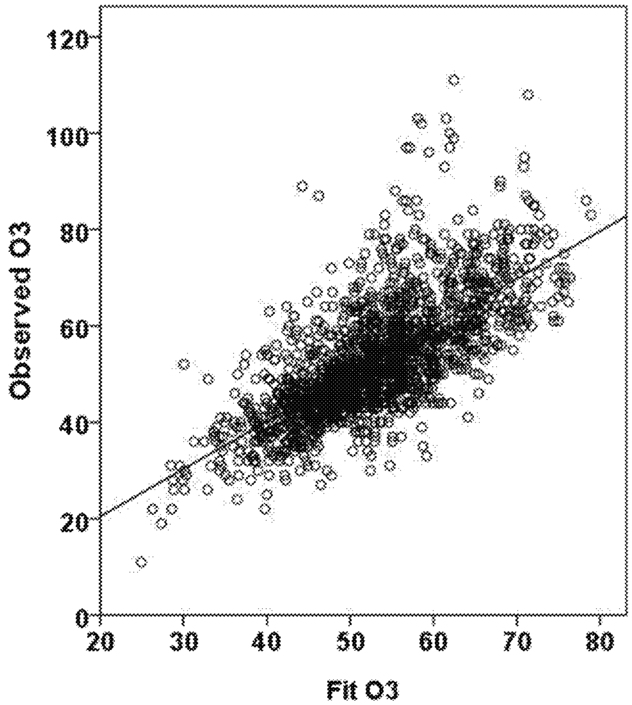


Figure A2. Observed vs Modeled O_3 (ppb) for the hours of 10am-2pm. The slope is 0.99, intercept is 0.7 and the R^2 is 0.49.

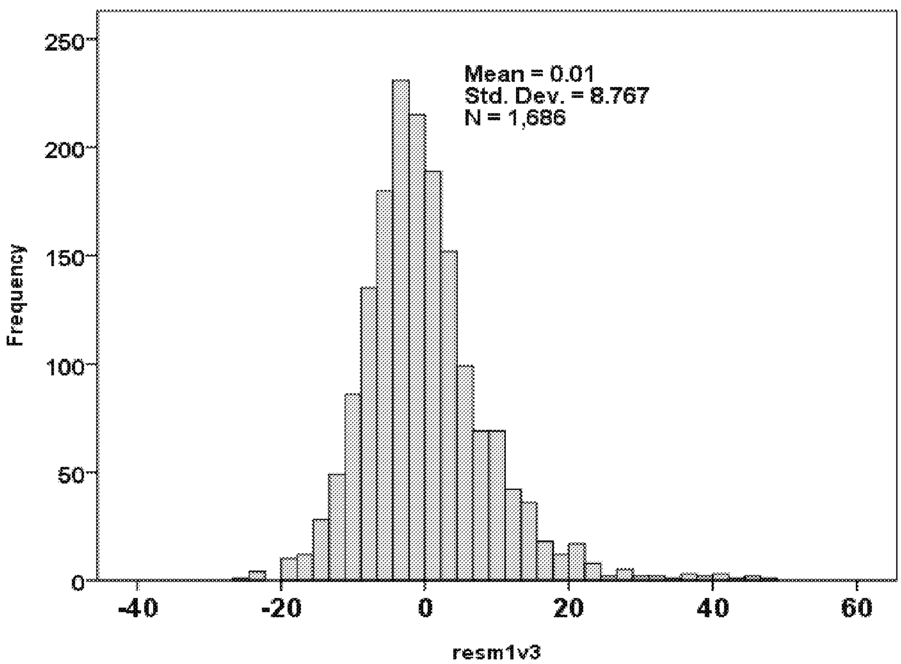


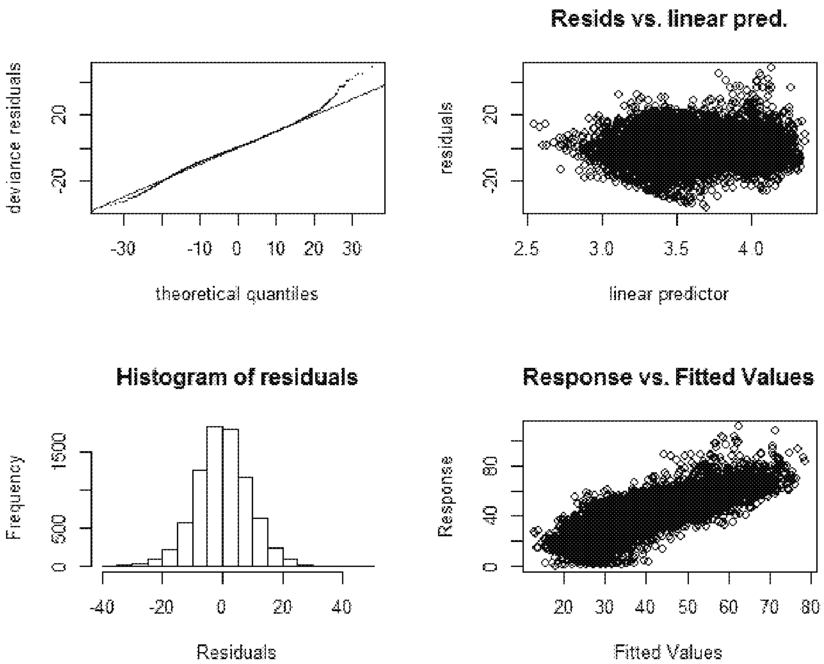
Figure A3. Histogram of residuals for modeled O_3 (ppb) for the hours of 10am-2pm.

Next we include a summary of the GAM results from R. This shows the model form and equation and the P values for each individual predictor. Also shown are the standard graphical outputs from the GAM "summary" function. The last item is a copy of the actual GAM code that we developed to run the model with the El Paso hourly O_3 and meteorology data.

Summary of model results from GAM function in R:

```
> AIC(m1)
[1] 58800.42
> summary(m1)

Family: gaussian
Link function: log

Formula:
O3 ~ s(V1, bs = "cr", k = 4) + s(V2, bs = "cr", k = k2[1]) +
    s(V4, bs = "cr", k = k2[4]) + s(V5, bs = "cr", k = k2[4]) +
    s(V6, bs = "cr", k = k2[4]) + s(V7, bs = "cr", k = k2[4]) +
    s(V8, bs = "cr", k = k2[4]) + s(V9, bs = "cr", k = k2[4]) +
    s(V10, bs = "cr", k = k2[4]) + s(V11, bs = "cr", k = k2[4]) +
    s(V12, bs = "cr", k = k2[4]) + s(V13, bs = "cr", k = k2[4]) +
    s(V14, bs = "cr", k = k2[4]) + s(V15, bs = "cr", k = k2[4]) +
    s(V3, bs = "cr", k = k2[4]) + Month + WD1017 + TR16Q

Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  5.005e+00  1.847e-01  27.106  < 2e-16 ***
Month       -2.008e-01  2.859e-02  -7.025 2.31e-12 ***
WD1017       2.970e-04  5.196e-05   5.715 1.13e-08 ***
TR16QNW     -1.429e-01  1.584e-02  -9.024  < 2e-16 ***
TR16QSE     -1.466e-01  1.281e-02 -11.439  < 2e-16 ***
TR16QSW     -1.485e-01  1.531e-02  -9.700  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
        edf Ref.df        F  p-value
s(V1)  3.000  3.000  100.251  < 2e-16 ***
s(V2)  8.931  8.997   21.633  < 2e-16 ***
s(V4)  8.939  8.999 1084.239  < 2e-16 ***
```

```
s(V5)  8.437  8.901   21.541  < 2e-16 ***
s(V6)  8.461  8.882   14.497  < 2e-16 ***
s(V7)  8.204  8.803    6.207 2.77e-08 ***
s(V8)  8.876  8.989    6.664 9.65e-10 ***
s(V9)  8.835  8.987   20.665  < 2e-16 ***
s(V10) 7.862  8.671    7.615 5.72e-11 ***
s(V11) 8.999  9.000    9.890 3.23e-15 ***
s(V12) 6.226  7.553    7.134 5.85e-09 ***
s(V13) 8.939  8.997   12.951  < 2e-16 ***
s(V14) 8.384  8.856    7.483 6.28e-11 ***
s(V15) 8.561  8.915   22.055  < 2e-16 ***
s(V3)  8.628  8.936   20.628  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) =  0.639   Deviance explained = 64.5%
GCV = 87.291  Scale est. = 85.911    n = 8047
```



GAM code for analysis of El Paso hourly data using R software:

```
rm(list = ls(all = TRUE))
library(mgcv)

# these are the parameters for the spline fit (Knots)
#Use K=4 for HOURLY--ACTUALLY USING K=10
k1 = c(10,10,10,10,10,10,10,10,10,10)
k2 = c(10,10,10,10,10,10,10,10,10,10)

### Next line reads in data file from my computer.
### You will need to configure the path correctly for your computer.
dat = read.csv("C:/ElPaso/GAM/ELP_hr_JJ_final4.csv",header=T)

#Model below is best for hourly data:
#For hourly model, I found best fit is with the gaussian family and log link
#Running this code on 10 months of hourly data, takes about one hour on my laptop.

m1 = gam(O3~s(V1,bs="cr",k=4)+s(V2,bs="cr",k=k2[1])+s(V4,bs="cr",k=k2[4])+
        s(V5,bs="cr",k=k2[4])+s(V6,bs="cr",k=k2[4])+s(V7,bs="cr",k=k2[4])+
        s(V8,bs="cr",k=k2[4])+s(V9,bs="cr",k=k2[4])+s(V10,bs="cr",k=k2[4])+
        s(V11,bs="cr",k=k2[4])+s(V12,bs="cr",k=k2[4])+s(V13,bs="cr",k=k2[4])+
        s(V14,bs="cr",k=k2[4])+s(V15,bs="cr",k=k2[4])+s(V3,bs="cr",k=k2[4])+
        Month+WD1017+TR16Q,family="gaussian"(link="log"),
        data=dat,na.action=na.exclude,sefit=true)

#Basic model checks below
AIC(m1)
summary(m1)
plot(m1)

#Calculate fit and residuals
resm1 = residuals(m1,type="response")
fitm1 = fitted.values(m1)

#THE NEXT LINE CAN BE USED TO GET PREDICTIONS FROM INPUT VARIABLES.
se3=predict.gam(m1,dat, type="response", se.fit=TRUE)

#Bind the output and ptint it out.
datnew = cbind(dat,fitm1,resm1,se3$fit)

### Next line is for my computer
### You will need to configure the path correctly for your computer.
write.csv(datnew,"C:/Dans/ElPaso/Results_NEWFILENAME.csv")
```